On numbers related to Catalan numbers

Carl Pomerance
Mathematics Department, Dartmouth College,
Hanover, NH 03755, USA
carl.pomerance@dartmouth.edu

September 13, 2013

Abstract

We study some numbers which have a definition similar to the Catalan numbers.

1 Introduction

Let $C(n)$ denote the nth Catalan number, defined as

$$C(n) = \frac{1}{n+1}\binom{2n}{n}.$$

There are numerous combinatorial applications of the Catalan numbers, see Stanley [5], and from any one of these we see that $C(n)$ is an integer, being the solution to a counting problem. From a number-theoretic perspective, one can ask for a direct proof of the integrality of $C(n)$. Of course, this is not difficult, perhaps the easiest way to see it is via the identity

$$C(n) = \binom{2n}{n} - \binom{2n}{n-1}.$$

But this raises the further issue of why the binomial coefficients are themselves integers, which is perhaps not immediately obvious from the formula

$$\binom{m}{k} = \frac{m!}{k!(m-k)!}.$$

For a prime p and a positive integer m, let $v_p(m)$ denote the number of factors of p in the prime factorization of m. For example, $v_2(10) = 1$, $v_2(11) = 0$, and $v_2(12) = 2$. This function can be extended to positive rational numbers via $v_p(a/b) = v_p(a) - v_p(b)$. Thus, a/b is integral if and only if $v_p(a/b) \geq 0$ for all primes p.

It is easy to see that

$$v_p(m!) = \sum_{k \leq m} v_p(k) = \sum_{k \leq m} \sum_{1 \leq j \leq v_p(k)} 1 = \sum_{j \geq 1} \sum_{\substack{k \leq m \\ p^j | k}} 1 = \sum_{j \geq 1} \left\lfloor \frac{m}{p^j} \right\rfloor,$$

since $\lfloor m/p^j \rfloor$ is the number of multiples of p^j in $\{1, 2, \ldots, m\}$. This result, sometimes referred to as Legendre's formula, is of course well known in elementary number theory, as well as the almost trivial inequality

$$\lfloor x + y \rfloor \geq \lfloor x \rfloor + \lfloor y \rfloor \tag{1}$$

for all real numbers x, y. From these two results it is immediate that $\binom{m}{k}$ is an integer, since for each prime p we have

$$v_p\left(\binom{m}{k}\right) = \sum_{j \geq 0} \left(\left\lfloor \frac{m}{p^j} \right\rfloor - \left\lfloor \frac{k}{p^j} \right\rfloor - \left\lfloor \frac{m-k}{p^j} \right\rfloor \right) \geq \sum_{j \geq 0} 0 = 0. \tag{2}$$

There is another consequence of this line of thinking. For a real number x let $\{x\} = x - \lfloor x \rfloor$, the fractional part of x. The inequality (1) can be improved to an equation:

$$\lfloor x + y \rfloor - \lfloor x \rfloor - \lfloor y \rfloor = \{x\} + \{y\} - \{x + y\} = \begin{cases} 1, & \text{if } \{x\} + \{y\} \geq 1, \\ 0, & \text{if } \{x\} + \{y\} < 1. \end{cases}$$

So by (2), $v_p(\binom{m}{k})$ is exactly the number of values of j such that $\{k/p^j\} + \{(m-k)/p^j\} \geq 1$. Let's write k and $m - k$ in the base p, so that

$$k = a_0 + a_1 p + \ldots, \quad m - k = b_0 + b_1 p + \ldots,$$

where the "digits" a_i, b_i are integers in the range 0 to $p - 1$. For $j \geq 1$,

$$\left\{ \frac{k}{p^j} \right\} = \frac{a_0 + a_1 p + \cdots + a_{j-1} p^{j-1}}{p^j}, \quad \left\{ \frac{m-k}{p^j} \right\} = \frac{b_0 + b_1 p + \cdots + b_{j-1} p^{j-1}}{p^j},$$

and so we see that $\{k/p^j\} + \{(m-k)/p^j\} \geq 1$ if and only if in the addition of k and $m-k$ in the base p there is a carry into place j caused by the earlier digits.

We thus have the remarkable result of Kummer from 1852: *For each prime p and integers $0 \leq k \leq m$, $v_p\left(\binom{m}{k}\right)$ is the number of carries in the addition $k + (m-k) = m$ when done in the base p.*

An immediate consequence of Kummer's theorem is that the nth Catalan number $C(n)$ is an integer. Indeed, say p is a prime and $v_p(n+1) = j$. Then the least significant j digits in base p of $n+1$ are all 0, so the least significant j digits in base p of n are all $p-1$. Thus, in the addition $n + n = 2n$ performed in the base p, we have j carries from the least significant j digits, and perhaps some other carries as well. So we have $v_p(\binom{2n}{n}) \geq j$. Since this is true for all primes p, we have $n + 1 \mid \binom{2n}{n}$ and so $C(n) = \binom{2n}{n}/(n+1)$ is an integer.

This of course is not the easiest approach to seeing that $C(n)$ is an integer, but the proof "has legs", meaning that it can be used to see some related interesting results. One might wonder if other numbers near to $n+1$ can divide $\binom{2n}{n}$. In particular, one could ask if the numbers $\binom{2n}{n}/(n+k)$ are integral for a fixed value of $k \neq 1$. We show that in no case are they always integral, but for $k > 0$, they usually are.

Theorem 1. *For each integer $k \neq 1$ there are infinitely many positive integers n with $n + k \nmid \binom{2n}{n}$.*

Thus, the case $k = 1$ of Catalan numbers is indeed special. However, the set of numbers n satisfying the condition of Theorem 1 when $k \geq 2$ is rather sparse. To measure how dense or sparse a set S of positive integers is, let $S(x)$ denote the number of members of S in $[1, x]$. Then the "asymptotic density" of S is $\lim_{x\to\infty} S(x)/x$ if this limit exists. In general, the limsup gives the upper density of S and the liminf the lower asymptotic density. For example, the set of odd numbers has asymptotic density $\frac{1}{2}$, the set of prime numbers has asymptotic density 0, and the set of numbers which have an even number of decimal digits does not have an asymptotic density.

Theorem 2. *For each positive integer k, the set of positive integers n with*

$$(n+1)(n+2)\cdots(n+k) \;\Big|\; \binom{2n}{n}$$

has asymptotic density 1.

So it is common for $n+k$ to divide $\binom{2n}{n}$, but what about $n-k$?

Theorem 3. *For each non-negative integer k the set of integers $n > k$ with $n-k \mid \binom{2n}{n}$ is infinite, but has upper asymptotic density smaller than $\frac{1}{3}$.*

Theorems 1 and 3 in the case $k=0$ might be compared with Ulas [6, Theorems 3.2, 3.4].

For other problems and results concerning divisibility properties of binomial coefficients, the reader is referred to [1, 4].

2 Proofs

There is a quick proof of Theorem 1. First assume that $k \geq 2$. Let p be a prime factor of k and let $n = p^j - k$ where j is large enough so that $n > 0$. Then n in base p has at most j digits and the least significant digit of n is 0. Hence there are at most $j-1$ carries when adding n to n and hence $n + k = p^j \nmid \binom{2n}{n}$. For $k \leq 0$, let $p > 2|k|$ be an odd prime number. For $n = p + |k|$, we have that there are no carries when n is added to itself in the base p, so that $p \nmid \binom{2n}{n}$. But $p = n + k$, so we are done.

The proof of Theorem 2 is a bit more difficult. We begin with a lemma.

Lemma 1. *For each prime p and all real numbers $x \geq 2$, the number of integers $1 \leq n \leq x$ with $p \nmid \binom{2n}{n}$ is at most $px^{1-\log(3/2)/\log p}$.*

Proof. If $p \nmid \binom{2n}{n}$, then by Kummer's theorem, every base-p digit of n is smaller than $p/2$. (In particular, this implies that $\binom{2n}{n}$ is always divisible by 2.) Let $D = \lfloor 1 + \log x/\log p \rfloor$, so that if $1 \leq n \leq x$ is an integer, then n has at most D base-p digits. If we restrict these digits so that they are all smaller than $p/2$, we thus have $\lceil p/2 \rceil$ choices in each place. Thus, the number of choices for n is at most $\lceil p/2 \rceil^D$. It remains to note that $\lceil p/2 \rceil \leq \frac{2}{3}p$ so that

$$\lceil p/2 \rceil^D \leq \left(\tfrac{2}{3}p\right)^D < px\left(\tfrac{2}{3}\right)^{\log x/\log p} = px^{1-\log(3/2)/\log p}. \tag{3}$$

□

Since the quotient $px^{1-\log(3/2)/\log p}/x$ tends to 0 as $x \to \infty$ (with p fixed), Lemma 1 shows that for a given prime p, we usually have $p \mid \binom{2n}{n}$, in that the set of such numbers n has asymptotic density 1. We now wish to strengthen this lemma to show that p^j usually divides $\binom{2n}{n}$ for fairly large values of j.

Lemma 2. *Let p be an arbitrary prime, let $x \ge p$ be a real number, and let $D = \lfloor 1 + \log x/\log p \rfloor$. The number of integers $1 \le n \le x$ with $v_p(\binom{2n}{n}) \le D/(5\log D)$ is at most $3px^{1-1/(5\log p)}$.*

Proof. The calculation here is similar to the one in probability where you compute the chance that a coin flipped D times lands heads fairly frequently. We consider the number of assignments of D base-p digits where all but at most $B := \lfloor D/(5\log D)\rfloor$ of them are smaller than $p/2$. Since there are $\lceil p/2 \rceil$ integers in $[0, p/2)$, the number of assignments is at most

$$\sum_{j=0}^{B} \binom{D}{j} \big(\lceil p/2\rceil\big)^{D-j} \big(p - \lceil p/2 \rceil\big)^{j} = \sum_{j=0}^{B} \binom{D}{J} \big(\lceil p/2\rceil\big)^{D} \left(\frac{p - \lceil p/2\rceil}{\lceil p/2 \rceil}\right)^{j}$$
$$\le \big(\lceil p/2\rceil\big)^{D} \sum_{j=0}^{B} \binom{D}{j}.$$

A crude estimation using $D \ge 2$ gets us

$$\sum_{j=0}^{B} \binom{D}{j} \le \sum_{j=0}^{B} D^j < 2D^B,$$

and using the definition of B, we have

$$2D^B = 2e^{B\log D} \le 2e^{D/5} \le 2e^{1/5}x^{1/(5\log p)}.$$

Using the inequality in (3), we conclude that the number of $n \le x$ with $v_p(\binom{2n}{n}) \le B$ is at most

$$2\big(\lceil p/2\rceil\big)^{D} D^B \le 2e^{1/5}px^{1-\log(3/2)/\log p + 1/(5\log p)}.$$

Since $2e^{1/5} < 3$ and $\log(3/2) > 2/5$, the lemma follows at once. ∎

We are now ready to prove Theorem 2. Fix a value of $k \ge 1$. For a positive integer n and a prime p, let

$$v := v_p\big((n+1)(n+2)\cdots(n+k)\big).$$

First note that for each prime $p \ge 2k$,

$$v_p\left(\binom{2n}{n}\right) \ge v. \tag{4}$$

Indeed, since p divides at most one of $n+1, n+2, \ldots, n+k$, if $v > 0$, there is a unique positive integer $j \le k$ with $v_p(n+j) = v$. And since the v least significant digits of $n+j$ in base p are 0, the v least significant digits of n are at least $p-k \ge p/2$. So by Kummer's theorem, $p^v \mid \binom{2n}{n}$. (Note that this argument is essentially a reprise of the proof given in the Introduction that Catalan numbers are integers.)

It remains to show that (4) holds for all primes $p < 2k$, and for "most" integers n. Let $x \ge 2k$ be a real number and assume that we are considering values of $n \le x$. With D as in Lemma 2, we consider two cases: $v \le D/(5\log D)$ and $v > D/(5\log D)$. In the first case, by Lemma 2, the number of n with $v_p(\binom{2n}{n}) < v$ is at most $3px^{1-1/(5\log p)}$. Summing this expression for primes $p < 2k$ gives a quantity smaller than $6k^2 x^{1-1/(5\log(2k))}$. Divided by x, this expression tends to 0 as $x \to \infty$, so we are left with the second case when $v > D/(5\log D)$. We shall show in this case that there are very few values of n to consider, regardless if (4) holds. Let

$$v' = \max_{1\le i\le k} v_p(n+i),$$

and note that (as in the proof Legendre's formula)

$$v - v' < \frac{k}{p} + \frac{k}{p^2} + \cdots = \frac{k}{p-1} \le k.$$

For each value of i in $[1,k]$ we consider those $n \le x$ with $p^{v'} \mid n+i$. The number of them is at most $(x+i)/p^{v'} \le 2x/p^{v'}$. Since there are k possibilities for i, the number of choices for $n \le x$ is at most

$$k\frac{2x}{p^{v'}} \le 2k\frac{x}{p^{v-k}} < 2kp^k\frac{x}{p^{D/(5\log D)}} \le 2kp^k x^{1-1/(5\log D)}.$$

Now $\log D \le \log(1+\log x/\log 2)$, call this expression $L(x)$. Summing for $p < 2k$, we get that the number of choices for n is at most $k^2(2k)^k x^{1-1/(5L(x))}$. When divided by x, this too goes to 0 as $x \to \infty$, which completes the proof of Theorem 2.

We remark that the proof allows for k to also tend to infinity, provided it does not do so too quickly in comparison with x.

We now proceed to the proof of Theorem 3. Suppose that $k \ge 0$, n is in $(k,x]$, $x > 2k^2$, and p is the largest prime factor of $n-k$. Suppose too that $p > \sqrt{2x}$. If $n-k = cp$, then $n = cp+k$ and this is the base-p representation of n. Since $x > 2k^2$, the digit k of n is smaller than $\sqrt{x/2} < \frac{1}{2}p$. Further, $c \leq x/p < \sqrt{x/2} < \frac{1}{2}p$. Thus, there are no carries when adding n to itself in base p, so that $p \nmid \binom{2n}{n}$, and so $n - k \nmid \binom{2n}{n}$.

We now show that there are many values of n in $(k, x]$ such that $n - k$ has a prime factor $p > \sqrt{2x}$, in fact more than $\frac{2}{3}x$ of them. For each prime p satisfying this inequality and $x > 2k^2$, we count numbers n in $(k, x]$ with $p \mid n - k$, and this is $\lfloor (x-k)/p \rfloor \geq \lfloor x/p \rfloor - 1$. No choice of n corresponds to two different values of p, since their product would be too large to have $n \leq x$. Thus, the number of choices for n with $k < n \leq x$ and $n - k \nmid \binom{2n}{n}$ is at least

$$\sum_{\sqrt{2x} < p \leq x} \big(\lfloor x/p \rfloor - 1 \big).$$

Using $\lfloor x/p \rfloor > x/p - 1$, this count exceeds

$$x \sum_{\sqrt{2x} < p \leq x} \frac{1}{p} - 2\pi(x),$$

where $\pi(x)$ denotes the number of primes in $[1, x]$. Euler proved long ago in 1737 that the sum of the reciprocals of the primes diverges to infinity like the double-log function, and using a more modern estimation (such as the theorem of Mertens from 1874), we have

$$\sum_{\sqrt{2x} < p \leq x} \frac{1}{p} = \log\log x - \log\log(\sqrt{2x}) + E(x),$$

where $E(x) \to 0$ as $x \to \infty$, see [2, 3]. The difference of double logs simplifies to

$$\log 2 + \log\left(\frac{\log x}{\log(2x)}\right),$$

which tends to $\log 2$ as $x \to \infty$. As mentioned before, the set of primes has asymptotic density 0, in fact $\pi(x)/x$ goes to 0 like $1/\log x$ by the prime number theorem. Putting these thoughts together, it follows that for each $\epsilon > 0$ and x sufficiently large, there are more than $(\log 2 - \epsilon)x$ values of n with $k < n \leq x$ and $n - k \nmid \binom{2n}{n}$. Since $\log 2 > 0.6931 > \frac{2}{3}$, the second part of Theorem 3 follows.

To complete the proof, we wish to show there are infinitely many values of n where $n - k \mid \binom{2n}{n}$. We leave the few details for the reader, but using

Kummer's theorem, we have that if $n = pq + k$ where p, q are primes with $k < p$ and $\frac{3}{2}p < q < 2p$, then $n - k \mid \binom{2n}{n}$. The number of such numbers $n \le x$ is greater than a positive constant times $x/(\log x)^2$.

It is not clear if the numbers n with $n - k \mid \binom{2n}{n}$ comprise a set of positive density. Perhaps some numerical investigations are warranted; the case $k = 0$ is especially attractive.

3 Odd rows of Pascal's triangle

One might wonder what the fuss is about the middle entry $\binom{2n}{n}$ in the $2n$th row of Pascal's triangle. What about odd-numbered rows, where there are the twin peaks $\binom{2n+1}{n} = \binom{2n+1}{n+1}$? It is easy to prove corresponding results on divisibility here. For example, for $k \ge 2$, $(n+2)(n+3)\dots(n+k)$ usually divides $\binom{2n+1}{n}$.

Looking at just the case $k = 2$, we have a near miss for $n + 2$ always dividing $\binom{2n+1}{n}$. In fact, unless $n + 2$ is a power of 2, it divides, and even in this case, it divides $2\binom{2n+1}{n}$. Since there is a tie for the maximum entry in this row of Pascal's triangle, it makes sense to include both of them, and as mentioned we always have

$$n + 2 \;\Big|\; 2\binom{2n+1}{n}.$$

Thus, we might wonder, like with Catalan numbers, if the integer

$$\frac{2}{n+2}\binom{2n+1}{n} \tag{5}$$

has combinatorial significance. Indeed it does. We know that the Catalan number $C(n)$ counts the number of paths from $(0, 0)$ to (n, n) that do not cross below the line $y = x$, and where each step of the path is one unit to the right or one unit up. The number in (5) is similar, but now we are counting paths from $(0, 0)$ to $(n, n + 1)$, a so-called ballot number. We have come full circle back to combinatorial considerations.

Acknowledgments. I wish to thank Sergi Elizalde, Florian Luca, and Paul Pollack for some helpful comments.

References

[1] A. Granville, "Arithmetic properties of binomial coefficients. I. Binomial coefficients modulo prime powers", in *Organic mathematics (Burnaby, BC, 1995)*, pp. 253-276, CMS Conf. Proc., 20, Amer. Math. Soc., Providence, RI, 1997.

[2] P. Pollack, *Not always buried deep, a second course in elementary number theory*, Amer. Math. Soc., Providence, RI, 2009.

[3] P. Pollack, "Euler and partial sums of the prime harmonic series", *Elem. Math.*, to appear.

[4] J. W. Sander, "A story of binomial coefficients and primes", *Amer. Math. Monthly* **102** (1995), 802-807.

[5] R. P. Stanley, *Enumerative combinatorics, vol. 2*, Cambridge University Press, Cambridge, 1999.

[6] M. Ulas, "A note on Erdős–Straus and Erdős–Graham divisibility problems (with an appendix by Andrzej Schinzel)", *Int. J. Number Theory* **9** (2013), 583-599.